Deep Learning and the Steam Engine Before Thermodynamics: A Historical Comparison

Abstract

Modern deep learning systems are extraordinarily successful yet remain poorly understood theoretically. This situation bears a striking resemblance to the development of steam engines prior to the emergence of thermodynamics. Early steam engineers built increasingly capable machines through empirical experimentation, scaling, and practical ingenuity, long before the scientific principles governing efficiency and optimality were understood. This note explores the analogy and suggests that contemporary AI may occupy a similarly pre-theoretical stage.

The Pre-Thermodynamic Steam Age

By the late eighteenth and early nineteenth centuries steam engines had already transformed mining, manufacturing, and transportation. Engineers such as James Watt produced machines of immense practical value despite possessing little understanding of energy conservation, entropy, or theoretical efficiency limits.

Progress occurred primarily through:

- empirical experimentation,
- scaling to larger machines,
- engineering intuition,
- incremental improvements,
- extensive trial and error.

Design choices were often justified pragmatically rather than theoretically. If a larger boiler or higher operating pressure improved performance, it was adopted even if the underlying reasons remained obscure.

Only later did the work of Carnot, Clausius, Kelvin and others establish thermodynamics, revealing deep principles governing all heat engines. Suddenly questions that had previously seemed empirical acquired precise theoretical answers:

- What is the maximum possible efficiency?
- Which engine cycles are optimal?
- Where are losses occurring?

- Which improvements are fundamentally impossible?

The development of thermodynamics transformed engine design from largely empirical craft into mature engineering science.

Contemporary Deep Learning

Modern deep learning exhibits many similar characteristics.

Neural networks already perform remarkable tasks including language understanding, image recognition, scientific discovery and code generation. However, their development remains heavily empirical.

Progress is often driven by:

- larger models,
- larger datasets,
- increased computation,
- architectural experimentation,
- extensive hyperparameter tuning.

Numerous practical heuristics are known:

- initialization schemes,
- normalization methods,
- residual connections,
- attention mechanisms,
- learning rate schedules,
- regularization techniques.

Yet these methods frequently resemble engineering recipes more than consequences of a unified theory.

The field still lacks generally accepted answers to fundamental questions:

- Why do highly overparameterized networks generalize?
- What is the minimal computational structure required for intelligence?
- What are the fundamental limits of learning efficiency?
- What architectural principles are essential rather than incidental?
- How much routing complexity is actually required?

The Compute Expansion Analogy

Early steam engineers often compensated for inefficiency by building larger machines and burning more fuel.

Modern AI frequently compensates for incomplete understanding by scaling:

- more parameters,
- more GPUs,
- more training data,
- longer training runs.

This strategy has proven remarkably successful. However, it may conceal substantial inefficiencies.

Future researchers may discover that many contemporary architectures expend enormous computational resources exploring state spaces that are vastly larger than necessary.

Routing, Gating and the Search for Deeper Principles

One candidate for a future organizing principle is routing.

Conventional ReLU networks possess an enormous theoretical routing state space generated by independently switching nonlinear units. Training may spend considerable effort learning to coordinate these switches into coherent and stable computational pathways.

If only a tiny subset of the available routing configurations is actually useful, current networks may contain large amounts of unnecessary computational freedom.

Future theories could reveal:

- optimal routing constraints,
- minimal state spaces,
- preferred computational geometries,
- fundamental efficiency limits.

Such discoveries would parallel the role played by thermodynamics in revealing the optimal operation of heat engines.

Possible Future "Thermodynamics of Intelligence"

A mature theory of intelligent computation may ultimately synthesize ideas from:

- information theory,
- dynamical systems,
- statistical mechanics,
- control theory,
- geometry,
- network science.

Such a framework might provide principled answers to questions currently addressed largely through experimentation.

It may establish quantities analogous to:

- energy,
- entropy,
- efficiency,
- irreversibility,
- conservation laws.

These concepts could define hard limits on learning, representation and computation.

Conclusion

Deep learning may presently occupy a historical position analogous to steam engineering before thermodynamics: highly successful, economically transformative, yet still lacking a comprehensive theoretical foundation.

If this analogy is correct, future theoretical advances may yield efficiency improvements far exceeding those obtainable through scaling alone. Future AI systems may eventually regard today's trillion-parameter models much as modern engineers regard early steam engines: historically important, ingenious, and surprisingly inefficient.

This analogy should not be taken too literally. Deep learning already rests upon substantial mathematics from optimization, statistics, information theory and dynamical systems. Nevertheless, the comparison captures the sense that the field's practical successes currently outpace its theoretical understanding.